

C Exhibit



By Rebecca Hellweil | Apr 6, 2021 12:00pm EDT

YouTube says it's better at removing videos that violate its rules, but those rules are in flux

YouTube's newest content moderation stat, briefly explained.

TECHNOLOGY SOCIAL MEDIA YOUTUBE

Vox

We have a request

Vox's journalism is free, because we believe that everyone deserves to understand issues that impact them. Will you give today to help keep Vox free for all?

Yes, I'll Give

x

and that's what my team works day in and day out to try to do."

YouTube shared this new information as politicians and users have grown increasingly concerned about how technology companies are moderating their platforms amid an "infodemic" of **Covid-19 misinformation**, and following the **insurrection at the US Capitol** and a presidential election cycle last year that was marked by **conspiracy theories**.

At the same time, YouTube's stats on violative content bolster a narrative some YouTube executives have **promoted** in the past: that its systems generally do a good job of catching bad content, and that overall, the problem of nefarious videos on its site is **comparatively small**. YouTube also said on Tuesday that it's able to take down 94 percent of content that breaks its rules with automated flagging systems, and that the large majority of those videos are caught before they get 10 views. Overall, YouTube claims it's removed more than 83 million videos since it started releasing enforcement transparency reports three years ago.

"We have a large denominator, meaning we have lots of content," CEO Susan Wojcicki **told Recode** back in 2019. "When we look at it, what all the news and the concerns and the stories have been about this fractional 1 percent."

But the numbers YouTube released on Tuesday have limitations. Here's how it calculated them: YouTube samples a number of views, meaning instances in which a user looks at a particular video (YouTube did not release the number of videos that factored into this statistic). Then, YouTube looks at the videos getting those views and sends them to its content reviewers. They study all the videos and figure out which ones violate the company's rules, allowing YouTube to produce an estimated percentage rate of views that happened on "violative videos."

Keep in mind that YouTube's own reviewers — not independent auditors — decide what counts as a violation of YouTube's guidelines. While Facebook last year committed to **an independent audit** of its community standards enforcement metrics, Flannery O'Connor said on Monday that the video platform had yet to



By Rebecca Hellweil | Apr 6, 2021 12:00pm EDT

YouTube says it's better at removing videos that violate its rules, but those rules are in flux

YouTube's newest content moderation stat, briefly explained.

TECHNOLOGY SOCIAL MEDIA YOUTUBE

Yes, I'll Give

Vox's journalism is free, because we believe that everyone deserves to understand issues that impact them. Will you give today to help keep Vox free for all?

We have a request

Vox

x

and that's what my team works day in and day out to try to do."

YouTube shared this new information as politicians and users have grown increasingly concerned about how technology companies are moderating their platforms amid an "infodemic" of **Covid-19 misinformation**, and following the **insurrection at the US Capitol** and a presidential election cycle last year that was marked by **conspiracy theories**.

At the same time, YouTube's stats on violative content bolster a narrative some YouTube executives have **promoted** in the past: that its systems generally do a good job of catching bad content, and that overall, the problem of nefarious videos on its site is **comparatively small**. YouTube also said on Tuesday that it's able to take down 94 percent of content that breaks its rules with automated flagging systems, and that the large majority of those videos are caught before they get 10 views. Overall, YouTube claims it's removed more than 83 million videos since it started releasing enforcement transparency reports three years ago.

"We have a large denominator, meaning we have lots of content," CEO Susan Wojcicki **told Recode** back in 2019. "When we look at it, what all the news and the concerns and the stories have been about this fractional 1 percent."

But the numbers YouTube released on Tuesday have limitations. Here's how it calculated them: YouTube samples a number of views, meaning instances in which a user looks at a particular video (YouTube did not release the number of videos that factored into this statistic). Then, YouTube looks at the videos getting those views and sends them to its content reviewers. They study all the videos and figure out which ones violate the company's rules, allowing YouTube to produce an estimated percentage rate of views that happened on "violative videos."

Keep in mind that YouTube's own reviewers — not independent auditors — decide what counts as a violation of YouTube's guidelines. While Facebook last year committed to **an independent audit** of its community standards enforcement metrics, Flannery O'Connor said on Monday that the video platform had yet to